1957

The Multiple Choice Rorschach: A Comparison of Two Scoring Systems in Estimating Psychiatric Aide Efficiency as Rated By Supervisors

Edward O. Treesh
Loyola University Chicago

THE MULTIPLE CHOICE RORSCHACH: A COMPARISON OF TWO SCORING SYSTEMS IN ESTIMATING PSYCHIATRIC AIDE EFFICIENCY AS RATED BY SUPERVISORS

by

Edward O. Treesh

A Thesis Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

June

1957

LIFE

Edward O. Treesh was born in Warsaw, Indiana, April 25, 1922.

He was graduated from Warsaw High School, Warsaw, Indiana, June, 1940, and from Indiana University, June, 1949, with the degree of Bachelor of Arts.

He began his graduate studies at Loyola University, September, 1949, and continued through June, 1951. From September, 1951, through August, 1955, he was employed as a staff psychologist at the Larned State Hospital, Larned, Kansas. In September, 1955, he enrolled at Purdue University, Lafayette, Indiana, where he is now engaged in graduate studies.

ACKNOWLEDGMENTS

The investigator wishes to acknowledge the invaluable assistance of Mr. Paul L. Reed, Personnel Director, Larned State Hospital, Larned, Kansas whose cooperation made this study possible.

The PGT scoring system was made possible through the cooperation of Dr. Stephen H. Pratt and Mr. Gordon A. Gardiner, members of the Department of Clinical Psychology, Larned State Hospital, Larned, Kansas.

The committee members, Dr. Vincent V. Herr, Chairman, Dr. Frank J. Kobler, and Dr. Edmund P. Marx are to be recognized for their constructive criticisms.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

PURPOSE OF THE STUDY

This study was designed as the first of a series in the development of a weighted scoring system for the Multiple Choice Test (MCT), one test utilized for screening applicants for the positions of psychiatric aides or hospital attendants.

Utilization of the data of a battery of tests, Kuder Preference Record, Cornell Index, Otis Employment Test, Multiple Choice Test, by a personnel manager in conjunction with past employment history and personal interview had resulted in a marked reduction of ward personnel turn over. Previously the hospital where the study was conducted was reported to have had the greatest turn over rate of the state institutions. At the time of the study the hospital had one of the lowest rates of employee turn over. The personnel manager attributed a part of this success to the interpretations of the one "projective" test, MCT, included in the battery of tests administered to applicants for ward personnel positions.

Clinical psychologists who administered the testing program had observed responses to the MCT believed to have predictability of employee satisfaction or dissatisfaction on the job; however there had been no opportunity for experimental tests of the tentative formulations. The scoring of the MCT was believed to be modifiable so that discrimination could be

increased.

This study was designed to compare the discrimination of the Harrower-Erickson scoring system and a revised system developed by two members of the clinical psychology staff in conjunction with the experimenter. The revised system will be referred to as the PGT system, a combination of the first letter of each of the psychologist's last names, Pratt, Gardiner, and Treesh. The test of the scoring systems was to differentiate between a group of employees rated as efficient as contrasted with another group of employees rated as less efficient.

Hypotheses to be investigated in this experimental study were as follows:

(1) The "high" efficiency group of employees will tend to score fewer "poor" responses as a group than the "low" efficiency employees when the test is scored with the original Harrower-Erickson scoring system.

(2) The "high" efficiency group of employees will tend to respond to more "positive" responses as a group than the "low" efficiency group when the test is scored with the PGT system.

(3) The PGT system will tend to discriminate between the groups with greater accuracy than the original system proposed by Harrower.

CHAPTER II

REVIEW OF RELATED RESEARCH

The original MCT was an outgrowth of the earlier Group Rorschach. The need for a screening instrument that could be administered to large numbers of individuals at the same time and scored by psychologically unsophisticated personnel was noted by Harrower (12, p. 117) as the motivating factor for the development of the First Multiple Choice Record Blank (12, p. 118). The Amplified Multiple Choice Test (12, p. 222) was introduced after the original MCT had been in use in an attempt to provide revisions suggested by extensive investigation during this period of time.

The original MCT blank consisted of lists of ten alternative responses for each of the Rorschach plates. The testee was instructed to underline the response which in his opinion was the best description of the blot or any of its parts. Included at the end of each list was a space where the testee could write in a response other than the alternatives listed above.

Modifications of the first MCT resulted in the Amplified Multiple Choice Test. There was no space provided where the testee could write something other than the above, and three groups of ten alternative responses for each plate were listed. Within each of these three groups the testee was requested to underline the response that was the best description of the blot

or any of its parts. Under this procedure, the testee underlined thirty responses instead of ten as was requested in the original MCT.

Scoring the original MCT consisted of noting the "poor" responses checked by the testee. "Poor" had been defined by Harrower as responses occurring more frequently in various types of psychopathological records. "Good" had been defined as responses occurring more frequently in "normal" records. Responses (12, p. 222-227) were assigned numerals for ease of scoring. "Good" were assigned the numerals one, two, three, four, and five. "Poor" were assigned the numerals six, seven, eight, nine, and ten. The quantitative score consisted of the total number of "poor" responses underlined by the testee, that is the number of responses to which the numbers six, seven, eight, nine, and ten were assigned. "Good" responses were not taken into consideration in the quantitative scoring. The score consisted only of those responses selected by the testee that Harrower had included from individual records of persons with various types of psychopathology.

To score the amplified MCT, the examiner noted the total number of "poor" responses and calculated a percentage of the total responses that were "poor". Number ten responses were each counted as two "poor" responses. Responses assigned numbers of six and seven were each counted as one-half "poor" answer. Responses assigned numbers eight or nine were each counted one "poor" response.

The underlying assumption of the test was that those persons most likely to respond with certain kinds of responses when responding freely in the Rorschach test would tend to pick such responses when confronted with such

responses in a multiple choice situation.

Harrower (11, 12) discussed her initial experiments with the original MCT and reported that only six to sixteen percent of the normal groups scored four or more poor responses. While at the other extreme seventy three to seventy nine per cent of the institutionalized patients and neuropsychiatric cases scored four or more poor responses. Intermediate between these extremes were students referred by psychiatrists and a sample of prisoners. Harrower (12, p. 128) has provided protocols obtained from the large number of subjects in her initial experimental groups.

One of the largest samples reported in the literature was composed of 3,150 psychotic patients, 883 attendants, 172 professional adults, and one hundred teen age girl scouts. Wittman[1] (32) utilized the original MCT in this study. If the scores of her groups were ranked from the poorest to the best the following list would result: schizophrenic behavior disorders, constitutional behavior reactions, affective behavior disorders, organic behavior reactions, patients without psychosis, paranoid behavior reactions, and extramural controls. Wittman suggested that the MCT measured a type of inner adjustment which might be labeled personality integration. She believed that the instrument might measure type rather than degree of adjustment.

There are numerous studies that have been reported. These studies were carried out in the following settings or with the following groups: applicants for jobs (1); student nurses (2); teachers (3); high-school

---

1 Dr. Wittman-Huffman

pupils (4, 5, 8); psychiatric screening (19); psychiatric patients (6, 17); college students (12, p. 247, 21, 22, 27); patients with spinal cord injuries (20); Armed Forces personnel (12, p. 161-206, 13, 28, 30, 31); and employee adjustment or efficiency (12, p. 240, 33).

One of the criticisms pointed out by various investigators was that utilization of only a quantitative score usually resulted in overlap of groups or failure to discriminate between groups. Overlap of different groups was noted by Wittson, Hunt, and Older (31) and Due, Wright, and Wright (12, p. 170) who reported that the utilization of a cutting score as the point of differentiation of groups resulted in too great an overlap in groups to warrant the use of the instrument by other than trained Rorschach personnel. Springer (28) was not able to differentiate between chronic offenders to Naval discipline and those sailors making a good adjustment. Winfield (30) used the test with women in the Marines but concluded that the test was measuring something other than it was reported to have been measuring. Jensen and Rotter (13) found that forty five per cent of their officers and thirty six per cent of the officer candidates had scores which fell in the pathological range and concluded that the test was unsuitable for these groups. Berg (2) experienced similar difficulty in attempting to discriminate between successful and unsuccessful student nurses.

Another criticism indicated by various investigators was that the scoring system resulted in significant differences between the sexes. This criticism was observed particularly in those studies dealing with adolescents or with college students (5, 8, 16).

That the original scoring system needed revision has been suggested by numerous investigators. Singer (24) in particular has indicated that the psychopathological significance of percepts utilized in multiple choice projective tests should be determined empirically. Harrower (12, p. 234) indicated that the response items had been chosen not by experimental techniques but a priori due to pressure of time.

Following an item analysis study, Malamud and Malamud (17) indicated that uniform scoring of only those answers falling into "poor" categories was insufficiently refined for optimal discrimination. They suggested that responses, even in the same category, might require increased weights on some cards, decreased or even no weight on other cards.

A series of "weighted" scoring systems were developed by Osborne, Sanders, and Green (22) for use in the prediction of academic achievement within a college setting. Their results indicated several types of weighted Rorschach responses which, when combined with scholastic aptitude scores, yielded appreciably better predictions of college marks than those based solely on scholastic aptitude tests. Their scoring systems were superior to the systems proposed by Harrower.

More recently Spoerl (27) has suggested that "category scoring" might result in greater discrimination. She suggested that the percentage of "poor" answers for those individuals who scored predominantly in categories one through five be compared to those individuals who scored predominantly in categories six through ten.

In their critique Smith and George (26) have noted the lack of

correlation between individual Rorschachs and MCT and suggested that the MCT is subject to conscious and deliberate faking as sexual and pathological responses are readily identifiable to normal subjects.

Cox (12, p. 240) has suggested that studies to determine the discrimination power of items for each use to which the MCT was employed was perhaps the most reasonable approach to the utilization of this screening device. Employees should be divided into high and low rating groups on the basis of criteria applicable to the particular setting. An item analysis might then be conducted to discover those items that differentiate.

Although there have been numerous criticisms of the MCT, there have been studies reported to indicate that the test is of value when utilized by trained Rorschach personnel. Wittman (32) reported the utilization of the test as a differential tool in a hospital setting. Due, Wright, and Wright (12, p. 108, 171) have indicated the application of interpretative principle and the utilization of the MCT in differential diagnosis.

A Rorschach Ranking Test was developed by Eysenck (9) as a modification of the original procedure. Four neurotic and five normal responses were included for each blot and were selected from the original list of Harrower. Subjects were asked to rank the responses in order of applicability. The best score would be thirty as the neurotic responses would be ranked as sixth, seventh, eighth, and ninth. The poorest score would be ten in which the neurotic responses would be given the first four ranks.

Kellman (14, 15) has recently developed a modified version of the MCT so that a profile similar to that obtainable from an individual Rorschach

is possible. There is no literature of research conducted with this modification at this time.

In a study to determine if psychological tests would contribute to an understanding of individual differences among psychiatric aides so as to aid in their selection and placement, Yerbury, Holzberg, and Alessi (33) utilized the Revised Beta Examination and the MCT. Their sample consisted of 113 unselected psychiatric aides employed after the study was undertaken. After six months on the job, the employees were divided into two groups of "good" and "poor" employees upon the basis of a rating scale devised by these authors for their study. To eliminate many doubtful employees from either group and to gain refinement, the authors reevaluated their classifications. A total of forty subjects remained from the original group of 62 "good" employees. Eight employees were eliminated from the original "poor" group reducing the number to forty three.

After refinement of the groups so that marginal employees were eliminated, they concluded that the use of the total number of underlined and checked poor anatomical responses and x-ray responses gave the most satisfactory results. The total weights on the Revised Beta were found to be more sensitive in discriminating between poor and good psychiatric aides than were the IQ scores. Utilizing both tests, they found that thirty two per cent of these very "poor" employees were identified while only five per cent of the very good employees were falsely identified. Yerbury, Holzberg, and Alessi (32) suggested that a more extensive battery of psychological tests should lead to even more fruitful results in the problem of selection and placement

of psychiatric aides.

In summary it can be reported that the chief criticism of the MCT has been directed toward the quantitative scoring systems proposed by the test originator. The reliability and validity of this test instrument have been studied with appropriate recommendations regarding modification of the scoring system. Various investigators have developed scoring systems applicable to the particular setting in which the instrument was to be utilized. Two other investigators have proposed modifications of the original procedure. The consensus of opinion is that the present scoring system does not discriminate in the manner reported by the test originator and requires modification before the instrument can be utilized in the manner proposed by the test originator. Other investigators have reported that the test may be of value when interpretative principles are utilized rather than the quantitative score.

CHAPTER III

PROCEDURE

Three clinical psychologists worked as a group eight months prior to the present investigation and rated the responses on the amplified MCT blank on a continuum ranging from a minus three through zero to a plus three (Appendix I). The psychologists utilized only unmarked MCT blanks in assigning weights to the various response items. "Stop" responses were included and represented items ordinarily considered to warrant careful study of the total responses in a record. A "stop" item included was "Bloody hands" to plate nine. "Personal reference" responses were also included. One "personal reference" item noted was "Part of my body" on plate one. The plus weights were assigned to those responses thought to be typical of a reasonably adequate adjustment. Minus weights were assigned to those items presumed to be indicative of psychological stress and/or immaturity. The zero weights were assigned to those items that usually would be considered neutral but the significance of which could best be determined by a careful analysis of the thirty responses chosen by the testee. This revised scoring system was not utilized however until present investigation was outlined.

A list of all personnel classified as psychiatric aides, psychiatric aide trainees, and hospital attendants was obtained from the personnel office. The list included the name of the employee, his birthdate, the original date

of employment, and a numerical rating given to the employee by his immediate supervisor. A routine procedure of the hospital was compliance with an annual rating of all employees on the appropriate forms provided by the state civil service commission.

The Employee's Evaluation Report (Appendix II) consisted of nine scales six of which were applicable to a class of personnel known as general employees who were not in supervisory positions. The six scales were designated: (1) quality of work; (2) quantity of work; (3) use of working time; (4) work attitudes; (5) adaptability; and (6) dependability. Employees were rated on each of these scales by their immediate supervisors. The rating then was made a part of their permanent employee records in the personnel office.

Employees could be rated by checking an adjective rating which had been assigned a numerical weight. There were four possible ratings on each scale. The total of the six numerical weights assigned to the individual scales could be converted into an adjective rating descriptive of the overall efficiency of the employee. The adjective ratings for the total score were: excellent, very good, good, fair, and unsatisfactory.

At the time of the study there were three hundred employees who were classified as psychiatric aides, psychiatric aide trainees, or hospital attendants. Two hundred and thirty of these employees had been rated, and their ratings were made available to the experimenter.

Initial "high and low" efficiency groups were selected by assigning arbitrary cutting points to the numerical scale of ratings so that the majority of rated employees would be excluded from the experimental groups in

terms of their numerical rating. The initial high group consisted of sixty eight employees. The initial low group consisted of forty nine employees.

The initial high and low efficiency groups were then listed alphabetically and according to sex. The final experimental groups were selected by means of a table of random numbers (29) so that the sex distribution within each group would be representative of the original high and low groups. Thirty two people were assigned to each group.

The employee screening test files were then inspected to determine the extent to which members of the groups had been previously tested. Many of the employees who were selected for the experimental groups had taken the amplified MCT when the testing program was initiated at the hospital. Others had taken the test prior to their employment but since the inauguration of the testing program. A small minority of the groups had not taken the test. They had been excluded from the program initially since an arbitrary age limit of fifty five had been established as the terminal age for testing. These few employees were contacted while they were on duty and asked to cooperate in the experiment that was being conducted by the investigator. They were not aware of the purpose of the study.

The amplified MCT had always been given as a self administering test. Blots were to be viewed in one position, and the employee was requested not to rotate the blots so that the procedure would duplicate the original administration proposed by Harrower.

After all subjects had completed the tests, the experimenter scored all records by both scoring systems, Harrower and PGT. This scoring was in-

dependently checked by one other individual to prevent scoring errors before the results were recorded. The means and standard deviations of age, experience in months, and numerical efficiency ratings for the low and high efficiency groups are presented in Table I. Scores on the MCT will be presented in Table II and Table III.

TABLE I

MEANS AND STANDARD DEVIATIONS OF AGE, EXPERIENCE IN MONTHS, AND NUMERICAL EFFICIENCY RATINGS FOR LOW AND HIGH EFFICIENCY GROUPS

	Low Efficiency (N=32)		High Efficiency (N=32)			
Variable	Mean	S.D.	Mean	S.D.	t	Level of Significance
Age	49.81	15.41	47.38	11.40	.72	—
Experience	30.97	25.89	37.28	24.90	.99	—
Rating	13.51	0.67	18.38	0.55	31.49	.001

The data in Table I may be interpreted to indicate that the experimental groups do not differ significantly with respect to the variables of age and experience. That the groups differ significantly with respect to the variable of efficiency rating may be inferred.

CHAPTER IV

RESULTS

Prior to the experiment, the investigator elected to employ a non-parametric statistical procedure as no assumptions concerning the nature of the distributions would be necessitated. At the same time there was no reasonable basis for an inference other than that the scores or ratings would be merely modes of ranking the experimental groups.

The statistical test selected was the Sum of Ranks (29, p. 434). This test was first proposed by Wilcoxon and has since been derived by other workers in the field such as Mann and Whitney and White. Hypotheses can be formulated which are essentially a test in difference of location of groups on a continuum. One-tailed or two-tailed tests of significance may be employed with this statistical procedure.

For each of the scoring systems, a single distribution of the sixty four observations was made and ranks from one to sixty four were assigned. Rank one was assigned to that score assumed to represent the more adequate adjustment in terms of the quantitative score on either system utilized. On the PGT continuum, rank one was assigned to that score of the greatest positive magnitude. Rank one was assigned to that score which represented the least number of "poor" responses on the H-E continuum. Ranks were then assigned to each of the scores on the separate continua so that the rank of

sixty four would be assigned to that score representing the greatest number of "poor" responses on the H-E continuum or to that score of the greatest negative magnitude on the PGT continuum. If ties were present, the tied observations were given the average of ranks they would otherwise occupy. The raw scores and their respective ranks for each of the scoring continua are reported in Table II and Table III.

The basic statistical formula employed was that reported by Walker and Lev (29, p. 434). However since there were numerous ties present in each set of sixty four ranks, appropriate correction factors were utilized. The correction factor for each continuum was calculated in the manner suggested by Edwards (7, p. 426). To employ the correction factors within the basic statistical formula reported by Walker and Lev, the investigator then followed the suggestions of Edwards (7, p. 429) and modified the denominator of the basic formula to permit inclusion of the correction factor.

This statistic, Sum of Ranks, is reported (29) to have a distribution which is approximately unit normal if the number of cases in each group is as large as eight or larger. In accordance with the procedure outlined by Walker and Lev, the present investigator utilized the values tabled (29, p. 456) to determine if the z_0 values were statistically significant.

For purposes of statistical tests of the experimental hypotheses, the investigator assigned rank one to the left end of the continuum of ranks for each respective scoring system. Rank sixty four then became the terminal rank on the right end of the continuum of ranks for each respective scoring system.

TABLE II

RAW SCORES AND ASSIGNED RANKS OF THE LOW EFFICIENCY GROUP FOR THE PGT AND H-E SCORING SYSTEMS

	PGT System				Harrower System	
Subject	Score	Stop	Personal Reference	Rank	Total number poor responses	Rank
1	- 9	0	0	37.5	14.5	47.0
2	+44	0	0	2.0	5.5	12.0
3	- 9	1	0	37.5	15.0	49.0
4	+15	1	0	16.0	6.5	15.5
5	-17	1	0	42.5	11.0	30.5
6	+13	0	1	17.0	7.0	19.0
7	- 7	0	0	33.0	11.5	33.0
8	-41	0	1	56.0	14.5	47.0
9	+ 9	0	0	18.0	7.5	21.5
10	-49	1	1	59.0	14.0	43.0
11	-56	0	0	61.0	31.0	62.0
12	-18	0	1	44.0	10.5	28.0
13	-17	0	0	42.5	19.5	55.5
14	-28	0	0	47.5	16.0	51.0
15	-34	0	0	51.0	19.5	55.5
16	-28	0	0	47.5	13.0	38.0
17	0	1	1	29.0	7.0	19.0
18	-59	0	0	62.0	24.0	59.0
19	+16	0	0	15.0	6.5	15.5
20	-40	1	1	55.0	24.0	59.0
21	-23	0	1	45.5	14.0	43.0
22	- 8	1	0	35.0	16.0	51.0
23	-36	2	1	53.0	12.0	34.5
24	- 4	4	1	31.0	9.0	24.5
25	-38	0	0	54.0	16.0	51.0
26	-48	1	0	58.0	22.5	57.0
27	+ 3	0	0	25.5	11.0	30.5
28	+26	0	0	9.0	4.0	5.0
29	+ 1	0	0	27.5	16.5	53.0
30	-50	0	2	60.0	17.5	54.0
31	+ 7	0	0	20.0	8.5	23.0
32	- 7	0	0	33.0	13.0	39.0
	Mean				Mean	
	-15.375				13.688	
	SD				SD	
	25.51				6.15	

TABLE III

RAW SCORES AND ASSIGNED RANKS OF THE HIGH EFFICIENCY GROUP FOR THE PGT AND H-E SCORING SYSTEMS

	PGT System				Harrower System	
Subject	Score	Stop	Personal Reference	Rank	Total number poor responses	Rank
1	+ 1	0	0	27.5	12.0	34.5
2	+32	0	0	5.0	3.5	3.5
3	+37	0	0	3.0	5.0	10.0
4	-33	0	1	49.0	24.0	59.0
5	+24	1	0	11.5	4.5	7.0
6	+ 8	0	1	19.0	7.5	21.5
7	-34	0	0	51.0	25.5	61.0
8	+30	0	0	6.5	3.5	3.5
9	+ 4	1	0	23.5	10.0	26.5
10	+ 4	0	0	23.5	6.5	15.5
11	-13	0	0	40.5	12.5	36.5
12	-13	2	0	40.5	13.0	39.0
13	+25	0	0	10.0	6.5	15.5
14	-47	1	1	57.0	14.0	43.0
15	+21	0	0	13.0	6.0	13.0
16	+20	0	0	14.0	4.5	7.0
17	+33	0	0	4.0	2.0	2.0
18	-62	0	0	63.0	44.0	64.0
19	- 7	0	1	33.0	12.5	36.5
20	- 1	0	1	30.0	14.0	43.0
21	- 9	0	0	37.5	11.0	30.5
22	+ 3	0	0	25.5	9.0	24.5
23	+24	0	0	11.5	5.0	10.0
24	-68	0	1	64.0	36.0	63.0
25	+64	0	0	1.0	1.0	1.0
26	-34	0	1	51.0	14.5	47.0
27	- 9	1	0	37.5	10.0	26.5
28	-23	0	0	45.5	14.0	43.0
29	+ 6	1	0	21.0	7.0	19.0
30	+28	0	0	8.0	4.5	7.0
31	+ 5	0	0	22.0	11.0	30.5
32	+30	0	0	6.5	5.0	10.0
	Mean				Mean	
	+ 1.438				11.219	
	SD				SD	
	30.07				9.45	

The experimental hypotheses were formulated in the manner now acceptable to experimental sciences. The first hypothesis to be tested was in regard to the original Harrower scoring system and was formulated as follows:

H_1: Location of the high efficiency group on the Harrower scoring dimension is the same as that of the low efficiency group.

H_2: Location of the high efficiency group on the Harrower scoring dimension is to the left of the low efficiency group.

Level of significance: .05

Decision rule: Reject H_1 if Z_o is equal to or less than -1.64.

Z_o = -2.51 H_1 is rejected at the five per cent level of significance.

The second experimental hypothesis was formulated in terms of the PGT system of scoring.

H_1: Location of the high efficiency group on the PGT scoring dimension is the same as that of the low efficiency group.

H_2: Location of the high efficiency group on the PGT scoring dimension is to the left of the low efficiency employee group.

Level of significance: .05

Decision rule: Reject H_1 if Z_o is equal to or less than -1.64.

Z_o = -2.48 H_1 is rejected at the .05 level of significance.

A third experimental hypothesis was to be tested with respect to the discriminatory efficiency of the two scoring systems. The experimenter believed that the PGT system would be more efficient in differentiating between the high and low efficiency groups. The very small difference in

efficiency between the two scoring systems was not sufficient to warrant a statistical test. The PGT system resulted in twenty one high and eleven low employees being ranked in the first thirty two on the continuum of ranks for this particular scoring system. On the Harrower system twenty high and twelve low were ranked in the first thirty two ranks of this scoring system.

Utilization of "stop" and "personal reference" as two additional discriminators did not result in finer quantitative discrimination between the groups. Twelve subjects within the high group selected "stop" or "personal reference" responses as compared with fourteen subjects with the low efficiency group.

CHAPTER V

DISCUSSION

The results of the experiment may be interpreted to indicate that the high and low efficiency employees may be differentiated as groups. However individuals within either group obtained ranks throughout the continuum of ranks. The highest and lowest ranks of the PGT scoring dimension were assigned to a member of the low group. Inspection of the ranks assigned to the Harrower continuum indicates that the highest and lowest ranks were obtained by members of the high group. Rank number four was assigned to a member of the low group.

Considerable overlap of the experimental groups was noted even though statistically significant results were obtained. Within the first thirty two ranks on the PGT continuum there were twenty one members of the high group and eleven members of the low group. Twenty members of the high group and twelve members of the low group composed the first thirty two ranks on the Harrower continuum.

In settings where a low selection ratio can be employed, that is, when the object is to select few individuals from a large number of applicants the test might be effectively employed in selection procedures. Such a selection procedure however would be dependent upon the utilization of an entirely quantitative approach to selection.

The combination of the quantitative score with a qualitative interpretation of the record has been utilized to increase the discrimination. The PGT system of scoring was an attempt to quantify some of the qualitative features that might go unobserved when a strictly quantitative approach was utilized. Results of the experiment may be interpreted to indicate that the PGT system did not increase the discrimination power of the test except negligibly.

The chief contribution of the PGT system would seem to be the differentiation between groups as expressed in numerical scores. The high group achieved a mean score of +1.438 as compared to the low group with a mean score of -15.375. Utilization of the PGT resulted in a greater range of scores than was possible with the Harrower system. The mean number of poor responses for the high group was 11.22 as compared to 13.69 with the low group.

Results of the experiment might be accounted for in terms of the distribution of sexes within the groups. The high group was composed for the most part of females as compared with the low group made up predominantly of males. In view of the economic conditions surrounding the hospital at the time of the study, the personnel manager believed that the distribution of sexes in the experimental groups reflected the distribution of sexes according to rating of efficiency. More emotionally mature women were assumed to accept the relatively low salary as compared with the more emotionally mature men who were not as likely to apply for work at the hospital. At the time of the study numerous male conscientious objectors were employed at the institution

and tended to offset this bias.

The experimental groups did not differ significantly with respect to the variables of age and experience; however they were significantly different with respect to efficiency rating. That the ratings were made by different raters could be an uncontrolled variable in this experiment. Since the ratings were a part of the employee's permanent record it was assumed that the raters made the ratings as adequately as they were able to do so. According to civil service regulations an employee could be dismissed or placed on probation as a result of the rating he received.

A combination of the strictly quantitative and a qualitative analysis of the responses appear to be the most realistic approach to the utilization of the MCT. Discrimination with the MCT is too uncertain to warrant its use except in a test battery selected for a particular screening situation.

CHAPTER VI

SUMMARY

This study was designed as the first in a series in the development of a weighted scoring system for the amplified Multiple Choice Test developed by Harrower. Three clinical psychologists working as a group weighted responses on the MCT and included "stop" and "personal reference" categories as proposed means of increasing the discrimination of the MCT.

Both scoring systems were tested in the discrimination of a group of thirty two high efficiency psychiatric ward personnel from a group of thirty two low efficiency psychiatric ward personnel. Statistical tests indicated that both scoring systems were capable of discriminating between the groups beyond the one per cent level of significance. There was no significant difference between the scoring systems in discriminating between the groups.

Inclusion of "stop" and "personal reference" categories to the PGT scoring key added no significant discrimination power.

Utilization of the MCT in the screening of employees when a small ratio is to be selected from large numbers of applicants is indicated. Wide variation of the scores for the individuals in the experimental groups of psychiatric ward personnel employed in the study was noted.

BIBLIOGRAPHY

1. Balinsky, B. The multiple choice group Rorschach test as a means of screening applicants for jobs. J. Psychol., 1945, 19, 203-208.

2. Berg, I. A. A study of success and failure among student nurses. J. Appl. Psychol., 1947, 389-396.

3. Blair, G. M. Personality adjustments of teachers as measured by the multiple choice Rorschach test. J. Educ. Res., 1946, 39, 652-657.

4. Blair, G. M. & Clark, R. W. Personality adjustments of ninth-grade pupils as measured by the multiple-choice Rorschach test and the California Test of Personality. J. Educ. Psychol., 1946, 37, 13-20.

5. Blair, G. M., & Hoehn, A. J. The use of the multiple choice test with adolescents. J. Educ. Res., 1947, 41, 297-304.

6. Challman, R. C. The validity of the Harrower-Erickson multiple choice test as a screening device. J. Psychol., 1945, 20, 41-48.

7. Edwards, A. L. Statistical Methods for the Behavioral Sciences. New York: Rinehart & Company, Inc., 1954.

8. Engle, T. L. The use of the Harrower-Erickson multiple choice (Rorschach) test in differentiating between well-adjusted and mal-adjusted high-school pupils. J. Educ. Psychol., 1946, 37, 550-556.

9. Eysenck, H. J. A comparative study of four screening tests for neurotics. Psychol. Bull., 1945, 42, (9) 659-662.

10. Harrower-Erickson, M. R. A multiple choice test for screening purposes (for use with Rorschach cards or slides). Psychosom. Med., 1942, 5, 331-341.

11. Harrower-Erickson, M. R. & Steiner, M. E. Large Scale Rorschach Techniques: a manual for the Group Rorschach and Multiple Choice Test. Springfield, Ill: C. C. Thomas, 1945.

12. Harrower, M. R., & Steiner, M. E. Large Scale Rorschach Techniques: a manual for the Group Rorschach and Multiple Choice Test. (2nd Ed.) Springfield, Ill: C. C. Thomas, 1951.

13. Jensen, M. B., & Rotter, J. B. The validity of the multiple choice Rorschach test in officer candidate selection. Psychol. Bull., 1945, 42, 182-185.

14. Kellman, S. A proposed revision of the multiple-choice Rorschach: theoretical and methodological problems. Rorschach Res. Exch., 1949, 13, 244.

15. Kellman, S. Tests of personality: Rorschach techniques. D. Multiple Choice Rorschach. In Weider, A. Contributions Toward Medical Psychology. New York: Ronald Press, 1953.

16. Lawshe, C. H., Jr., & Forster, M. H. Studies in projective techniques: I The reliability of a multiple choice group Rorschach test. J. Appl. Psychol., 1947, 31, 199-211.

17. Malamud, R. F., & Malamud, D. I., The validity of the amplified multiple choice Rorschach as a screening device. J. Consult. Psychol., 1945, 9, 224-227.

18. Malamud, R. F., & Malamud, D. I., The multiple choice Rorschach: a critical examination of its scoring system. J. Psychol., 1946, 21, 237-242.

19. Marmor, J., & Coville, W. J. A study of the Harrower-Erickson multiple choice test as a psychiatric screening device. In Killinger, G. G., Psychobiological Program of WSA.

20. Mosak, H. H. Performance on Harrower-Erickson multiple choice test of patients with spinal cord injuries. J. Consult. Psychol., 1951, 19, 346-349.

21. Osborne, R. T., & Sanders, W. B. Multiple choice responses of college achievers and non-achievers. Educ. and Psychol. Meas., 1949, 9, 685-691.

22. Osborne, R., Sanders, W. B., & Greene, J. E. The prediction of academic success by means of "weighted" Harrower-Rorschach responses. J. Clin. Psychol., 1950, 3, 253-258.

23. Pinneau, S. R. Differences between normals and various types of schizophrenics on a multiple-choice Rorschach test. 1953, Ph. D., U. Calif.

24. Singer, M. The validity of a multiple choice projective test in psychopathological screening. Psychol. Monogr., 1950, 64, (8) (no. 314), v, 40 p.

25. Sisk, H. L. A reply to Winfield's study of the Multiple choice Rorschach. J. Appl. Psychol., 1947, 31, 446-448.

26. Smith, S., & George, C. E. The Harrower multiple choice Rorschach: A critique. J. Proj. Tech., 1954, 18, 507-509.

27. Spoerl, D. T. Category-scoring of the multiple choice Rorschach. J. Soc. Psychol., 1953, 38, 287-291.

28. Springer, N. N. The validity of the multiple choice group Rorschach test in the screening of Naval personnel. J. Gen. Psychol., 1946, 35, 27-32.

29. Walker, H. M. & Lev. J. Statistical Inference. New York: Henry Holt, 1953.

30. Winfield, M. C. The use of the Harrower-Erickson multiple choice Rorschach test with a selected group of women in military service. J. Appl. Psychol., 1946, 481-487.

31. Wittson, C. L., Hunt, W. A. & Older, H. J. The use of the multiple choice group Rorschach test in military screening. J. Psychol., 1944, 17, 91-94.

32. Wittman, P. The use of the multiple choice Rorschach as a differential tool. J. Clin. Psychol., 1945, 1, 281-287.

33. Yerbury, E. C., Holzberg, J. D. & Alessi, S. L. Psychological tests in the selection and placement of psychiatric aides. Amer. J. Psychiat., 1951, 108, 91-97.

34. Zuckerman, S. B. A research suggestion in large scale Rorschach. J. Consult. Psychol., 1948, 12, 300-302.

APPENDIX I

HARROWER MULTIPLE CHOICE TEST

(For use with Rorschach cards or slides)

NAME:

AGE:

DATE:

OCCUPATION:

INSTRUCTIONS

You are going to see ten inkblot pictures one after another.

Begin by taking a good look at Inkblot 1 and see if it, or any part of it, reminds you of anything or resembles something you have seen.

Then read through each of the three groups of answers for Inkblot 1 (A, B, C).

Now underline the one answer in Group A, the one answer in Group B, the one answer in Group C, which you think is the best description of that inkblot or any of its parts. You, therefore, underline three answers for Inkblot 1.

When you have done this, if you wish, you may put a check beside any other answer in any of the three groups which you also feel is a good description of the inkblot or any of its parts.

Then do exactly the same thing for each of the other inkblots.

	PGT System Qualitative	PGT System Score	Harrower System Score
INKBLOT 1			
A			
Underline one answer here.			
An army or navy emblem		+2	3
Crumbling cliffs		-1	9
A bat		0	2
Nothing at all		-3	10
Two people		+3	1
A pelvis		-1	3
An X-ray picture		-2	7
Pincers of a crab		0	4
A dirty mess		-3	9
Part of my body	Personal reference	-2	6
B			
Underline one answer here.			
A headless figure with arms up		0	1
Vertebra		-1	3
Tiny boxing gloves		0	4
Spilt ink		-2	9
Someone's insides		-2	6
Nothing at all		-3	10
A butterfly flying		+2	2
Lava		-1	9
A coat of arms		+2	3
An X-ray of the chest		-2	7
C			
Underline one answer here.			
A Halloween mask		0	3
Storm clouds		-2	7
A moth		+1	2
Two people on a merry-go-round		+3	1
A bell in the center		+2	3
An X-ray picture of the spine		-2	7
Animal heads on the sides		+1	4
The stomach		-2	6
Nothing at all		-3	10
Eyes glaring at me	Stop	-3	8

	PGT System Qualitative	PGT System Score	Harrower System Score
INKBLOT 2			
A			
Underline one answer here.			
A bug somebody stepped on		-3	8
Nothing at all		-3	10
Two scottie dogs		+2	2
Little faces on the sides		0	4
A bloody spinal column		-3	6
A white top		0	5
A bursting bomb		-2	8
Two elephants		+2	2
Two clowns		+3	1
Red and black ink		-2	9
B			
Underline one answer here.			
An animal skin		-1	7
Two bears rubbing noses		+1	2
Faces of Indians on the side		0	4
Blood		-3	8
Nothing at all		-3	10
A white lamp		0	5
An exploding firecracker		-2	8
A red butterfly		+2	3
Two people playing pat-a-cake		+3	1
Red and black splotches		-2	9
C			
Underline one answer here.			
Two witches		0	1
Black and red paint		-2	9
Bear's heads		+2	2
An empty hole	Stop	-1	9
Faces carved in stone		0	4
Lungs and blood		-3	6
A white sting ray		0	5
A little temple in the center		+1	4
Nothing at all		-3	10
An erupting volcano		-2	8

	PGT System Qualitative	PGT System Score	Harrower System Score
INKBLOT 3			
A			
Underline one answer here.			
Two birds fighting		-1	2
Meat in a butcher's shop		-2	8
Two men pulling something apart		0	1
Part of my body	Personal reference	-2	6
Just colored blots		-2	9
A colored butterfly		+2	3
Spots of blood and paint		-3	8
Monkeys hanging by their tails		+2	2
A red bow tie		+2	3
Nothing at all		-3	10
B			
Underline one answer here.			
A red brooch		+1	3
A person's insides		-2	6
Two cannibals		0	1
Donald Ducks		+1	2
Fire and smoke		-2	8
Nothing at all		-3	10
Spilt paint		-2	8
Two women quarreling		0	1
Blood and dirt		-3	9
Alphonse and Gaston--"after you"		+3	1
C			
Underline one answer here.			
Two birds' heads		+2	4
A bloody stomach		-3	6
Two waiters bowing		+3	1
An X-ray picture		-2	7
Dirty spots and bloody spots		-3	9
A colored hair ribbon		+2	3
Lipstick splotches		-1	8
Falling cats		-1	2
Nothing at all		-3	10
Fish swimming		+2	2

	PGT System Qualitative	PGT System Score	Harrower System Score
INKBLOT 4			
A			
Underline one answer here.			
Head of an animal		+1	4
Lungs and chest		-2	6
A nasty, dirty mess		-3	9
A pair of boots		+2	4
A burnt mass		-3	9
Nothing at all		-3	10
A giant in a fur coat		0	1
An animal skin		+1	3
A big gorilla		-1	2
An X-ray picture		-2	7
B			
Underline one answer here.			
A little flower on the top		+1	4
The spine		-1	6
Dirty water		-3	9
Charlie Chaplin's feet		+3	3
A nightmare		-3	9
A man sitting down		+2	1
A fur rug		+2	3
Two Scottie dogs		+2	2
A black smudge		-3	9
Nothing at all		-3	10
C			
Underline one answer here.			
Clouds		-1	7
A bat		-1	2
A man seen from below		+1	1
Nothing at all		-3	10
Something squashed		-3	9
A frightening picture		-3	9
A person's insides		-3	6
Two little snakes		+1	4
Big overshoes		+2	3
A cow's head		+1	3

	PGT System Qualitative	Score	Harrower System Score
INKBLOT 5			
A			
Underline one answer here.			
A bird's beak		0	4
Something squashed		-3	9
A ballet dancer		+2	1
Nothing at all		-3	10
A map		-1	7
Sugar tongs		0	4
A moth		+2	2
Shoulders		-1	6
Smoke		-2	7
A rabbit's head		+2	4
B			
Underline one answer here.			
A man's face		0	3
A black mess		-3	9
Two men with arms folded		0	1
An island		-1	7
A policeman	Stop	-3	1
A bird flying		+2	2
A pelvis		-2	6
Tar and soot		-2	9
Nutcrackers		0	4
Nothing at all		-3	10
C			
Underline one answer here.			
An alligator's head		0	4
Nothing at all		-3	10
A smashed body	Stop	-3	9
A fan dancer		0	1
An X-ray picture		-2	7
Legs		+1	4
A bat or butterfly		+3	2
Lungs and chest		-2	6
Black clouds		-2	8
A pair of pliers		0	4

	PGT System Qualitative	Score	Harrower System Score
INKBLOT 6			
A			
Underline one answer here.			
Two kings' heads with crowns		+2	3
An X-ray picture		-2	7
Parts of the body		0	6
A totem pole		0	3
A fur rug		+2	3
Mud and water		-3	9
A polished post		0	4
Nothing at all		-3	10
A turtle		+1	2
A landslide		-3	9
B			
Underline one answer here.			
A dragon-fly		+1	2
The spinal column		-1	6
A cat's whiskers		+1	4
Male and female organs		0	6
An animal skin		+2	3
Dirty water		-3	9
A sceptre		0	3
A snake's head		-2	4
Nothing at all		-3	10
A spattered mess		-3	9
C			
Underline one answer here.			
A butterfly at the top		+2	2
An X-ray of the spine		-2	7
Feathers at the top		+1	4
A bear skin		+3	3
A leaf		-1	7
A table leg		0	4
Nothing at all		-3	10
Gushing oil		-2	9
A little man		0	1
Part of the body		0	6

	PGT System Qualitative	Score	Harrower System Score
INKBLOT 7			
A			
Underline one answer here.			
Smoke		-2	7
Two women talking		+3	1
Parts of the body		-1	6
Animals		+2	2
Nothing at all		-3	10
A white chandelier	Stop	0	5
Burning fragments		-3	9
Lambs' tail		+1	4
An X-ray picture		-2	7
Bookends		+1	3
B			
Underline one answer here.			
Men's faces with big noses		+1	3
A butterfly at the bottom		0	2
Dirt from the gutter		-3	9
Scotties		+3	2
A pelvis		-2	6
Indians with feathered caps		+3	1
Nothing at all		-2	10
Clouds		-2	7
An X-ray of part of the body		-2	7
A necklace		+2	3
C			
Underline one answer here.			
Children playing		+3	1
The lower part of the body		-2	6
Fog or mist		-2	7
Nothing at all		-3	10
A squashed frog	Stop	-3	7
Statues		+1	3
A gray mess		-3	9
A moth		0	2
Dogs playing		+2	2
A white lamp		0	5

	PGT System Qualitative	PGT System Score	Harrower System Score
INKBLOT 8			
A			
Underline one answer here.			
An orange or pink butterfly		+3	3
Shoulders, lungs, and stomach		-3	6
Nothing at all		-3	10
Just colors		-2	9
An emblem		+3	3
A pretty flower		+3	3
Heaven and Hell	Stop	-3	8
Two blue cushions		-1	3
Two bears climbing		+3	2
Colored clouds		0	7
B			
Underline one answer here.			
Flowers and leaves		+2	3
An X-ray picture		-3	7
Colored blobs		-3	9
A horseshoe crab		0	3
Nothing at all		-3	10
Blue flags		+2	3
Two animals climbing		+3	2
A colored coat of arms		+3	3
Fire and ice	Stop	-3	8
Parts of the body		-3	6
C			
Underline one answer here.			
A Christmas tree		+3	3
A medical picture		-1	8
Frogs' heads		-2	2
Life and Death	Stop	-3	8
A mountain at the top		+2	3
A design for wallpaper		+3	3
Inside the mouth		-3	8
Two beavers walking on colored rocks		+3	2
Nothing at all		-3	10
Colored ink splashed on paper		-3	9

	PGT System Qualitative	Score	Harrower System Score
INKBLOT 9			
A			
Underline one answer here.			
Sea horses		+3	3
Just spilt paint		-2	9
Flowers		+3	3
Parts of the body		-3	6
Smoke and flames		-3	8
Deer or horns of deer		+2	2
Nothing at all		-3	10
Two witches		+2	1
Bloody clouds	Stop	-3	8
A candle		+1	3
B			
Underline one answer here.			
Nothing at all		-3	10
A pink jacket		0	3
Just colors		-3	9
Tropical plants		+2	3
The stomach and intestines		-3	6
A forest fire		-3	8
An animal's head on the side		+1	2
Two gnomes		+2	1
Bloody hands	Stop	-3	8
A fountain		+1	3
C			
Underline one answer here.			
A tropical flower or orchid		+2	3
Lobsters		+1	3
The insides of a person		-3	6
An explosion		-3	8
Men's faces on the sides		+1	3
Nothing at all		-3	10
Two Santa Clauses		+3	1
Storm clouds at sunset		-3	8
A violin		+3	3
Messy colors		-3	9

	PGT System Qualitative Score	Harrower System Score
INKBLOT 10		
A		
Underline one answer here.		
Two people	+3	1
Spilt paint	-2	8
A Chinese print	+3	3
An X-ray picture	-3	7
Just colored ink spots	-3	9
Spiders, caterpillars, insects	+2	2
Parts of my insides Stop	-3	6
A colored map of California	0	3
Nothing at all	-3	10
A flower garden	+3	3
B		
Underline one answer here.		
Undersea picture	+3	3
Two little dogs sitting up	+3	2
Stomach and intestines	-3	6
A lot of colors	-2	9
A medical picture	-2	8
A design for wallpaper	+3	3
A child's painting	+3	8
Two ladies holding hands	+3	1
Nothing at all	-3	10
Lots of animals running around	0	2
C		
Underline one answer here.		
A blue flower	+2	3
Colored ink	-3	8
A picture of spring or fall	+2	3
Parts of the body	-3	6
Just colors	-3	9
Octopus and crabs	0	2
Bones	-3	6
Coral and seaweed	+2	3
Flowers	+2	3
Nothing at all	-3	10

APPENDIX II

Adjective Rating

EMPLOYEE'S EVALUATION REPORT

Name _____ Class _____

Agency _____ Report Period _____ to _____

Raters: Read complete instructions on other side before making ratings.

QUALITY OF WORK. This refers to neatness, accuracy, thoroughness of detail and general efficiency in meeting standards of quality for the type of work involved.

Exceptionally good / Very good / Satisfactory / Poor worker
/4/ Superior worker /3/ Above average /2/ Average worker /1/ Unsatisfactory

QUANTITY OF WORK. This refers to the amount of work completed which is of an acceptable quality. Consider the amount of work accomplished in comparison to what may be expected in the position.

Exceptionally good / Very good / Satisfactory / Poor worker
/4/ Superior worker /3/ Above average /2/ Average worker /1/ Unsatisfactory

USE OF WORKING TIME. This refers to the tendency to abuse leave, time of arrival at and leaving work, wasting time on the job, general industry and attention to duty.

Exceptionally good / Very good / Satisfactory / Poor worker
/4/ Superior worker /3/ Above average /2/ Average worker /1/ Unsatisfactory

WORK ATTITUDES. This refers to willingness to follow directions, co-operation with supervisors, acceptance of constructive criticism, and interest in work.

Exceptionally good / Very good / Satisfactory / Poor worker
/4/ Superior worker /3/ Above average /2/ Average worker /1/ Unsatisfactory

ADAPTABILITY. This refers to ability with which employees learns, his flexibility, dexterity or resourcefulness in meeting and handling new and varied situations.

Exceptionally good / Very good / Satisfactory / Poor worker
/4/ Superior worker /3/ Above average /2/ Average worker /1/ Unsatisfactory

DEPENDABILITY. This refers to the amount of supervision required, reliability and manner in which employee follows instructions.

Exceptionally good / Very good / Satisfactory / Poor worker
/4/ Superior worker /3/ Above average /2/ Average worker /1/ Unsatisfactory

SUPERVISION. This refers to ability in selecting and training subordinates, organizing and directing work, and delegating authority and responsibility.

Exceptionally good	Very good	Satisfactory	Poor worker
/4/ Superior worker	/3/ Above average	/2/ Average worker	/1/ Unsatisfactory

LEADERSHIP. This refers to the ability to command respect and gain co-operation of subordinates.

Exceptionally good	Very good	Satisfactory	Poor worker
/4/ Superior worker	/3/ Above average	/2/ Average worker	/1/ Unsatisfactory

JUDGMENT AND IMPARTIALITY. This refers to the ability to be impartial with employees and the public and to make and stand by sound decisions.

Exceptionally good	Very good	Satisfactory	Poor worker
/4/ Superior worker	/3/ Above average	/2/ Average worker	/1/ Unsatisfactory

---

Rated by_____ Reviewed by_____

Date_____ Date_____

EMPLOYEE'S EVALUATION REPORT
(BACK SIDE)

I certify that I have seen and discussed the ratings on this evaluation report.

Employee Signature_____

Date_____

Comments:

Signature_____

I certify that to my knowledge the evaluation of this employee is correct as noted. Final rating is approved.

Signature of Appointing Authority_____

Date_____

GENERAL INSTRUCTIONS

Rate persons who supervise other employees on all nine items. Rate all other employees on the first six items, only. Read description of each item heading. Then read all four degrees of evaluation before making rating. Rate one person at a time against requirements of the job he is holding. Make an (X) in the box which represents your evaluation of his work. When rating is completed, add up the total of the boxes you have checked. Compare the total with the score below and mark final rating in box on other side (upper-right-hand corner).

General Employees	Adjective Rating	Supervisory Employees
22-24	Excellent	32-36
18-21	Very Good	26-31
13-17	Good	19-25
10-12	Fair	14-18
6-9	Unsatisfactory	9-13

APPROVAL SHEET

The thesis submitted by Edward O. Treesh has been read and approved by three members of the Department of Psychology.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

Apr. 13, 1951
Date

Vincent V. Hartl
Signature of Adviser